FEB 1 7 1964

SUMMARY OF ACTIVITIES AS A NATIONAL SCIENCE FOUNDATION SENIOR POSTDOCTORAL FELLOW

Institute of Genetics
University of Pavia, Pavia, Italy

Howard M. Cann, M.D.
Fellowship No. 52066

In the period July 5, 1962 to July 5, 1963, I worked and studied at the Institute of Genetics, University of Pavia, Italy, under the sponsorship of the National Science Foundation. I concentrated my efforts, for the most part, in the statistical aspects of human genetics. Although I was not enrolled in any formal courses, I did attend English language lecture courses given by N.T.J. Bailey (Stochastic Processes) and M. Kimura (Mathematical Population Genetics), as well as numerous seminars.

While at the Institute of Genetics, I worked on a number of projects which are summarized below:

CANN

Hemophilia Study

The mutational and segregational basis of hemophilia A and B is being studied. Special questionnaire forms designed to give pertinent genetical information about these 2 bleeding diseases have been sent to hematologists in many countries, and we have asked that they complete a form on each of their patients with hemophilia A or B. Each form asks for the birth date of the patient, of his mother and father, and of his 4 grandparents; for the total number of brothers and the number of brothers suffering from hemophilia; for the presence or absence of the carrier state in his sisters or maternal aunts; for the presence of hemophilia in maternal uncles or other relatives; for the age of onset of the manifest disease in the patient; and for the diagnosis (A or B).

If mutation frequency increases with age, or if mutations are accumulated in the gamete-producing organs with time, one would expect an increased transmission of mutant-bearing gametes with increasing age of the parents. As the age

of parents at the time of fertilization increases, one would expect a rise in the frequency of carriers of a mutated gene (i.e. heterozygous at the particular locus involved) among the offspring. With respect to hemophilia A and B, which are both transmitted via the X chromosome, one might expect that cases of mutational origin would, on the average, possess at birth mothers of significantly greater age than those cases which have arisen through segregation. The reasoning is similar for maternal grandparents in whom mutation is revealed by the birth of an affected grandson: these grandparents should possess mean ages, at birth of mothers of patients, in excess of those for maternal grandparents in whom the hemophilia gene has segregated from its normal allele. Further, if the sex-linked mutation has occurred in one of the maternal grandparents, on the average, these relatives should be older at birth of the mother of the case being considered than were the paternal grandparents at the birth of the patient's father, if mutation accumulates with age. We are particularly interested in the contribution of the maternal grandfather since the male gametes are continually produced during the reproductive years. In females, the gametes are already produced, probably by birth, and certainly by puberty.

Thus, we are looking for a maternal grandparental age effect in cases of hemophilia from families in which there exists no history of the disease prior to the generation of the patient, i.e. in those cases which conceivably could be of maternal grandparental mutational origin. If our hypothesis is correct, we will see that, on the average, maternal grandparents were older than were paternal grandparents (used as controls) at the births of the patients' parents in these cases and also older, on the average, than were paternal and maternal grandparents at births of parents of cases from families with a history of hemophilia prior to the patients' generations.

If mutation does accumulate with age of mothers, hemophiliacs from families with no other history of hemophilia (present or past generations) will possess mothers whose average age at their birth is higher than that of mothers of patients with affected sibs or maternal relatives, other conditions being equal.

To date, some 220 forms have been completed and returned:

Hematologists in	No. of Forms		
	Hemophilia A	Hemophilia B	Unspecified
Germany	64	23	2
Netherlands	3	4	
Sweden	16	2	
Italy	12	1	
Argentina	81	10	2

1 case of a disease which is not hemophilia has also been received

Eleven of these forms were rejected because of information not consistent with a diagnosis of hemophilia A or B.

Analysis of variance of mother's ages partitioned by presence or absence of a family history of hemophilia and by family (male sibship) size revealed no significant difference between the average age of the mothers at birth of non-familial hemophiliacs and that of the mothers at birth of patients with a positive family history. Partitioning mother's age at birth of the patient eliminates any bias from the preponderance of smaller size sibships in which the mothers might be expected to be younger than mothers of larger sibships. Even with the use of this partition there did not appear to be any significant difference among ages of mothers for the various sibship sizes encountered in this study. In general, mothers have been younger at birth of single non-familial cases than have been mothers at birth of patients whose sibs or maternal relatives are affected.

A maternal grandparental effect has not been demonstrated. The following comparisons have been made:

Families with no history of hemophilia in earlier generations (i.e. families with possibility that hemophilia is caused from grandparental mutation)

Age maternal G.F. - Age paternal G.F. = X_1

Age maternal G.M. - Age paternal G.M. = X_2

Age maternal Grandparents - Age paternal Grandparents = X_3

In all three comparisons, the differences between the age of maternal and paternal grandparent(s) did not significantly differ from zero, although for each comparison the paternal grandparent(s) was(were) older (on the average). Furthermore, each of the three age differences did not differ significantly from 1.2 years, the maximum age difference possible assuming that:

$$\bar{X}_s - \bar{X}_u = \frac{a6^2}{M_o + a\bar{X}_u}$$

M_o = mutation rate at beginning of reproductive years

$\bar{X}_s - \bar{X}_u$ will be maximum when $M_o = 0$

Thus:

$$\bar{X}_s - \bar{X}_u = \frac{a6^2}{a\bar{X}_u} = \frac{6^2}{\bar{X}_u} = 36/30 = 1.2$$

Where $\bar{X}_s$ = mean age of maternal grandparent in whom mutation arises (or of cases with mutant gene).

$\bar{X}_u$ = average age of control grandparent. This value = 30 years (approximate)

a = increase of mutant proportion of gametes per year.

6^2 = variance of generation time (or variance in age of grandparent at birth of mother affected. This value = 36 years for English mothers(Penrose)

It was also found that maternal grandfathers of cases from families with no previous history were slightly younger than maternal grandfathers from families with a previous hemophilia history.

Thus the behavior of the ages of grandparents or of mothers is not what is expected with the hypothesis of mutation accumulation. The expected effect may be demonstrated with more cases. In this type of study where the correct reporting of family history and of birth dates is so crucial, one wonders what the role of incomplete reporting plays in the present results.

These hemophilia data are also being analyzed with modern computer techniques of segregation analysis. All families with two or more male sibs are being used for the analysis. We are studying only hemophilia A since relatively few cases of hemophilia B were reported.

Country	Hemophilia A No. of Families S $\geq$ 2		
	Total	Simplex Families r = 1	Multiplex Families r > 1
Germany and The Netherlands	29	21	8
Sweden	9	3	6
Italy	3	1	2
Argentina	55	37	18
Total	96	62	34

Assuming segregation frequency p = $\frac{1}{2}$ (for males) we have estimated X (sporadic case frequency in all affected cases) on the basis of 1) single selection (π, probability of ascertainment, = 0) and 2) truncate selection (π = 1). Under single selection we estimate that X = .50 $\pm$ .13 The parameters of p, π and X have values which fit the data for all countries reporting. Under truncate selection, although X was estimated, the maximum likelihood score and variance for p indicate an X^2 not compatible with the assumed value of 0.5.

Our analysis is based only on the number of affected and unaffected sibs in each family. The information about maternal uncles (affected and unaffected) is not complete. It should be noted that most of the forms from Germany do not contain a question as to the number of affected male sibs; just the total number of sibs is listed. We have been assuming that the patient is the only affected if there is no mention of other affected sibs. However, this may prove to be an unwarranted assumption upon follow-up, which must be done. There appears to be a deficit of affected sibs in cases from Germany and Argentina, as indicated by simple hand tabulations in which the patient is subtracted from the number of affected in each family, and numbers of affected and unaffected siblings compared.

Effect of Selected Variables on Secondary Sex Ratio

Data from some 880,000 Italian birth certificates for liveborn children and some 20,000 birth certificates for stillborn children for the year 1960 have been tabulated and made available to us for analysis. The data recorded for each birth contains the sex of the child, the age of the maternal grandfather at birth of the mother, the age of the mother at birth of the child, the order of birth represented by this birth, and the geographic region of birth. The object of the analysis was to look for an effect of maternal grandfather's age on sex ratio in the population of births. If mutation accumulates in the gamete pool with time, and if gametes are actively produced during the reproductive ages only in males, sex-linked lethal recessive mutations should occur with increasing frequency as age increases in the male. Since a sex-linked lethal recessive mutation can only be expressed in the male grandchildren of the male in whom the mutation has occurred, it is more likely that the mutation will be passed from older grandparents to mothers. The result of a sex-linked lethal recessive mutation in a population of births should be a deficit of males (or an increase in aborted male fetuses or stillborn male children). If mutation does accumulate with age, one would expect that the male-female ratio, or the proportion of males in a population of newborns, would tend to decrease as the age of the maternal grandfather at birth of the mother increases. It is this effect which we sought in this study. We also investigated the effect of increasing mothers' age (at birth of the child), the birth order, and the geographic region of birth on the sex ratio of the population of 880,000 Italian liveborn children. Similar studies were carried out with the stillborn group.

The sex ratio for all Italian livebirths in 1960 (M/F sex ratio) was 1.0553. For stillbirths the sex ratio was 1.2274.

When the frequencies of male and female livebirths were distributed by the age of grandfathers at birth of the mothers, no heterogeneity of these data among the ages could be demonstrated by the chi square determination. Although no significant

regression of these frequencies on age exists, the regression coefficient indicated a positive trend for males, i.e. a slight but nonsignificant increase of frequency of male newborns with increasing age of maternal grandfathers. No heterogeneity among stillborn male and female frequencies distributed among grandfathers' ages could be demonstrated and a regression coefficient was found which indicated a negative (but nonsignificant) trend of male stillbirths with increasing age of grandfathers.

To determine whether the lack of heterogeneity among birth frequencies partitioned by sex and distributed by age of grandfathers was due to a peculiar distribution with respect to one of the other variables available for analysis, a similar distribution (sex frequency by grandfathers' ages) was made for each age of mother at birth of the child, for each order of birth (1 to 8 or more) and for each geographic region of birth. Still no heterogeneity could be found for stillborn or livebirth data distributions within each of the subclasses listed.

Investigations of a possible effect of age of mother at birth of the child and of order of birth on the sex ratio of livebirth and stillbirth populations failed to demonstrate any heterogeneity within distributions of sex frequencies for both variables. Further, no heterogeneity was found among the livebirth sex ratio distributed by geographic region of birth.

When the stillbirth population was partitioned by sex and distributed by geographic region, however, a highly significant heterogeneity was found among the sex ratios. Furthermore it was found that the male stillbirth frequencies showed a positive and significant regression on geographic region in a roughly north to south axis. When Sardegna and Sicilia were omitted from the calculation, the regression became more significant and more positive. The values given to the independent variable, the regions, were far from precise, being successive integers from 1 through 11 for the eleven regional classifications, progressing from north to south, used by the Italian Istituto Centrale Di Statistica. To improve the precision of the representation of the independent variable, the stillbirth rate per

1000 total births, computed from the data for each region, was used. The stillbirth rate was found to roughly increase from northern through southern Italian regions. Computation of weighted and unweighted regressions revealed in each instance a significantly positive regression of frequency of male stillbirths on increasing regional stillbirth rate. Socioeconomic conditions decline, in general, the further south one progresses in Italy, and this decline could explain the increase in stillbirth rate. However, it is difficult to implicate a deteriorating socioeconomic environment in altering the stillbirth sex ratio beyond such vague statements as male fetuses, term and near term, seem more susceptible to an unfavorable environment than do their female counterparts.

Agglutination of Human Red Blood Cells by Mouse Isoantisera

We have found agglutinating activity for human red blood cells in a number of mouse isoantisera. If this activity is antibody, the development and enhancement of agglutination by albumin or with trypsinized erythrocytes suggests that it is similar to incomplete antibody. Since all erythrocytes so far tested react with positive isoantisera, it appears that the agglutination is mediated by antigen other than A, B, O, the usual Rh components, M or N, - possibly an unknown erythrocyte antigen. We hope to screen red blood cells from many donors in an attempt to identify the antigen involved.

Should this activity be an immune response to mouse tissue antigens, the immunogen is probably an antigenic complex and is certainly not limited to H - 2 antigen. Genetic dissimilarity of the tissue donor and the animal being immunized for antiserum production seems to be important for strong human hemagglutinating activity. This relationship can be investigated further by testing antisera of various donor-host combinations.

Low agglutination titers in pooled, normal mouse serum makes it conceivable that the activity is not an immune response, being enhanced coincidentally with or indirectly

by isoimmunization. Testing of individual samples of normal serum from animals of a variety of strains may contribute to further characterization. We might consider that handling of mice may immunize them to human antigens, but the high titers of agglutinating activity in positive sera suggest an immunization process of greater magnitude than skin to skin contact.

Absorption of the agglutinating ability of isoantisera by human erythrocytes is not direct proof of antibody but is somewhat reassuring. By altering various factors (e.g. erythrocyte antigenic make up, number of cells used, number of absorptions) in future absorptions, it may be possible to demonstrate more meaningful differential absorptions. Finally, it will be important to determine if various mouse antigens can absorb the human hemagglutinating activity.

Effect of Month of Birth on Selected Variable for Conscripts

Recent studies have indicated the possibility of a relationship between the month of birth and the development of mental retardation. Knobloch and Pasamanick analyzed birth dates of individuals admitted to a state school for mentally retarded children in the period 1913-1948. These authors found that significantly more had been born in the months January, February and March.

In an attempt to elucidate this finding and explore related phenomena, we have undertaken a survey in which the effect of the month of birth on a number of variables is being determined. The study so far has dealt with variables drawn from the records of a group of conscripts from the rural areas in the Province of Parma. These records give information pertaining to the medical history, physical examination, socioeconomic status, education and vital statistics, including birth date, of each conscript for the period 1900-1910. The records have been abstracted, coded and punched onto I.B.M. cards.

The plan of the study was straightforward. A number of variables available on punched cards were selected and each variable was distributed by the months of birth of the conscripts. This task was performed using I.B.M. sorting and tabulating machin-

ery. A relationship between each variable and the month of birth was sought by determining if heterogeneity of frequencies among the birth months existed. Chi square was used to test for heterogeneity and was determined by the electronic computer for each variable. Finally using only those variables which showed heterogeneity among birth months, a sine curve (periodic regression) regression was calculated for each variable subclassification upon month of birth and the significance of the deviation of the regression curve from the mean determined by the analysis of variance. The electronic computer was employed for these calculations.

The variables selected from the conscripts' records were:

Variable	Subclassification	MXN Table
Military fitness	Accepted, 'Partial' Rejection Rejection (on medical basis)	3X12
Reasons for Rejection	Possible Mental Deficiency*, Others (*cerebropathies, epilepsy, mental diseases, deaf-mute)	2X12
Hair Color	Blond, Red, Brown, Black	4X12
General Pigmentation	Light Healthy, Pale, Dark Healthy	3X12
General Pigmentation	Light Healthy and Pale, Dark Healthy	2X12
General Pigmentation	Light Healthy, Dark Healthy	2X12
Occupation	Farmer, Non-Farmer	2X12
Literacy	Literate, Partially Literate and Illiterate	2X12
Literacy	Literate and Partially Literate, Illiterate	2X12
Geographic Zone of	Each of the 11 zones of residence at time of conscription	11 X 12

Heterogeneity of frequencies among birth months was found for the following variables: general pigmentation, literacy and geographic zone of residence. All other variables were distributed homogeneously among the 12 months.

The periodic regression sine curves calculated from the data for general pigmentation for recruits classified as Light Healthy and as Dark Healthy deviated signifi-

.candy from the general mean ($p < 0.05$). The greatest frequency of birth of "Light Healthy" recruits occurred in September. The frequency then declined to the lowest point which occurred in March and then began to increase. A reciprocal behavior of the sine curve was found for more heavily pigmented recruits, the highest frequency of births of these recruits occurring in March and April and the lowest frequency in September and October.

In the case of literacy, literate recruits were born with the highest frequency in the month of November, the months of lowest frequency being April and May. The periodic regression curve for literate recruits and for pooled partially literate and literate recruits deviated significantly from the grand mean.

The sine curves computed from frequencies of month of birth of recruits from the 4 geographic zones consisting of communes Colorno, Mezzani, Sorbolo, Torrile, Fornovo, Taro, Pellegrino, Varano, Melegari, Corniglio, Monchio delle Cor, Palanzano, Tissano, Felino and Sala Baganza deviated significantly from the means. The highest birth frequencies for recruits from Colorno, Mezzani, Sorbolo, Torrile, Felino and Sala Baganza occurred in October while the lowest occurred in April. For recruits from Corniglio, Monchio delle Cor, Palanzano, Tizzano, Fornovo Taro, Pellegrino and Varano Melegari, the highest birth frequencies occurred in the spring months while the lowest occurred in Autumn and early winter.

All sine curve computations were made with relative frequencies transformed to radians ($\arcsin \sqrt{p}$) so that the regression values are not biased by the monthly birth frequencies for all the recruits.

The periodic affect of month of birth on literacy and geographic zone of residence is probably related to socioeconomic factors, while the affect on general pigmentation may indicate the impact of physical or biological environment on the developing embryo and fetus.

Further studies of the affect exerted by the month of birth are indicated. It is known that isoantibody titers vary periodically in cattle and man. An investigation of the periodicity of isohemagglutinin titers in man with special emphasis on phenomena

possibly related to incompatibility (such variables as miscarriages, isoantibody titer at birth, occurrence of hemolytic disease of the newborn being studied) is particularly indicated and easily achieved. There are some indications that the sex ratio in man is affected by the month of birth. This could well be the result of socioeconomic bias introduced by the method of data collection. It will be worthwhile to explore in more detail the relationship between sex ratio and month of birth.

Segregation Analysis of Diabetes Mellitus, Juvenile Type

Using data collected for a genetic study of juvenile diabetes mellitus (Simpson, N.E.: Annals of Human Genetics (1962) 26, 1) a study of the genetic parameters involved was carried out. The segregation analysis techniques of Morton were employed with the electronic computer. The hypothesis of an autosomal recessive single locus inheritance was tested. The data contained a group of matings which were possibly intercrosses (unaffected parents, one or more affected offspring) and another group of matings which were possible backcrosses (one affected parent, one or more affected offspring). All of the families used were ascertained through one or more of the affected progeny. The data which listed number of probands per family permitted us to estimate a probability of ascertainment of approximately 0.29 from sibships with more than one affected issuing from a possible intercross. Using this probability as one fixed parameter and assuming that the proportion of sporadic cases (phenocopies, diagnostic errors) was 0, we estimated the segregation frequency from all the intercross families of sibship size 2 or more from probability equations involving these 2 parameters. A value of approximately 6.7 per cent was found (indicating, if correct, a penetrance of about 26 per cent).

Turning to possible backcross data we assumed truncate selection, since we knew genotypes of the parents (assuming recessive inheritance) and assumed again (with more confidence in this instance) that the proportion of sporadic cases was 0. Estimating the segregation frequency with these backcross families and using the sporadic proportion of 0 and a probability of ascertainment of 1 in the same probability equations used for the intercross data, we found a segregation frequency of about 15 per cent,

approximately twice the segregation frequency for the intercross data, a consistent result. Again, penetrance is about 26 per cent.

Independent determination of penetrance based on the cumulative distribution of the age of onset of all affected offspring from intercross families with sibships of size 2 or more revealed a value of 49 per cent. This value may be consistent with the penetrance estimates from segregation frequency because the weighting involved in this independent estimate was derived from frequency distributions of age at death for all siblings, affected and unaffected, and these distributions were constructed from small numbers (50 - 70 deaths).

Administrative Aspects of the Fellowship Program

My comments in this area are those of praise. All aspects of the administration of my Fellowship were well-coordinated to my needs and did not distract me from my Fellowship activities. I was particularly pleased with the handling and disposition of stipends and travel funds which went very smoothly and which caused an absolute minimum of anxiety on my part.

Conclusion

I wish to assure the National Science Foundation of the eminent suitability of the Institute of Genetics for training in human genetics for future Fellows. Professor Cavalli-Sforza, Director of the Institute (and my sponsor and teacher) has developed an outstanding unit for research and study in all aspects of human genetics. Activities progressing in cell culture, biochemical genetics and statistical and demographical human genetics are of the highest quality. The facilities are adequate for research in any one of a number of areas and include a well-equipped and smoothly running cell culture laboratory, biochemistry facilities and an Olivetti electronic computer. Stimulation and teaching abound at the Institute because of the efforts of Professor Cavalli and all his associates, and are further boosted by visiting lectures of excellent caliber. The demographic data available to the Institute for analysis are of sufficient quality and quantity to insure an educational and productive experience for anyone

who is interested in human genetics. These data, incidentally, are the result of painstaking preparation, careful collection and truly remarkable foresight.

The year which I spent at the Institute of Genetics was an enriching experience and has contributed to my career in academic Pediatrics and research and study in heritable disease.